DOCUMENT RESUME

ED 214 972 TM 820 150

AUTHOR Gephart, William J.
TITLE Evaluation: Past, Present and Future: Phi Delta Kappa Occasional Paper 17.
INSTITUTION Phi Delta Kappa, Bloomington, Ind. Center on Evaluation and Research.
PUB DATE [72]
NOTE 29p.; Small print, pages 4-5.
AVAILABLE FROM Phi Delta kappa, Box 789, Bloomington, Ind. 47402 ($1.50).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Decision Making; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Models
IDENTIFIERS Evaluative Dimensions

ABSTRACT

This synthesis of the evaluation process in education presents a matrix illustrating the dimension of a continuum of models in their degree of relationship to a decision made. A second dimension categorizes models by their described process components. Extant models are seen as basically the same in slightly different language, and their proliferation as making the task of learning about evaluation too complex. To complement models to each other and enhance evaluation, it is suggested models be labeled by their characteristics. In a continuum of models, evaluation's relationship to decision-making is given. Categories of describing several models are discussed by their general nature, problem-solving strategies, and the activities in the processes. Products or outcomes of model activities and assessment criteria are additional categories. The past and future of evaluation and various problems are surveyed. Stated needs are: changes to improve the conceptual base of the field, effective meaningfulness to other education elements, and the training of evaluation personnel. Also included are lists of references and related titles of occasion papers. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214972

PHI DELTA KAPPA
EIGHTH AND UNION BOX 789
BLOOMINGTON, INDIANA 47401

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PRESENTS AN

OCCASIONAL PAPER

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D. Klieger
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 150

OCCASIONAL PAPER 17

EVALUATION: PAST, PRESENT AND FUTURE

by

William J. Gephart
Phi Delta Kappa
Bloomington, Indiana

FROM

CEDR
CENTER ON EVALUATION DEVELOPMENT AND RESEARCH

PREFACE

Models.

We dress them in flowing silks and satins, in smart plaids and tweeds to display the latest fashions for the coming season. We mold them from plastic to discover and explain the most efficient aerodynamics for a new high speed mass transit design. Hobbyists assemble classic automobiles, ships, and military tanks from model kits and learn the intricate gears, belts, and bolts that make each work. Architects create scale-built towns to display their conception of the most liveable and space efficient of inner city rennovation projects. High school sophomores learn the circulatory and digestive systems from life size models of the human body that are cut away to colorfully display the body organs and skeleton.

Evaluation theorists too, make use of the model. Their diagrams--simple or complex--and verbal representations are devised to better explain the processes involved in an effective evaluation. In the past decade at least ten new models of the evaluation process have been described. As with nearly any model, these visual and verbal representations can aid the student or practitioner in understanding unknown or difficult concepts.

And yet, a growing wave of opinion is solidifying against the proliferation of evaluation models. Many would argue the addition of each new model serves little more than to boost the ego of the inventor and to confuse the concepts for the student.

In this Occasional Paper, William J. Gephart voices such a concern. But, more importantly, he suggests a viable solution. Instead of introducing yet another evaluation model, Dr. Gephart presents a concise synthesis of the many existing models. Whether this synthesis proves to be broadly acceptable remains to be seen. One thing is sure though, the direction pointed out in this paper must be taken if the field of evaluation is to solidify and advance.

INTRODUCTION

This paper focuses on evaluation today. Its objective is a synthesis of the many models of the evaluation process into the most parsimonious set possible. It seems totally ridiculous that we have a Tylerian model of evaluation, a Scriven model, a Stake model, a Stufflebeam model, a UCLA model, a Provus model, another Scriven model, a Fortune/Hutchinson model, a Hammond model, an accreditation model, and on and on.

This proliferation of models has two negative effects. It makes the task of learning about evaluation unnecessarily complex. At present the student of evaluation must learn the nature of each of these models. As will be argued later, many of them differ only in the vocabulary used to describe them; a factor which adds confusion to the learning task. Secondly, the identification of evaluation models by the name of their originator builds competing camps of evaluation theorists each striving to elevate "their" model. Growth in the field would be enhanced if we began to focus on the complementary nature of these models.

As a start in that direction, we could at least move to the labeling of the many models in another manner. Instead of the use of the names of their originators, it is proposed here that we label the models by using terms or phrases which describe their general nature or central characteristics. If that were done, we would refer to them as the Formative-Summative model instead of the Scriven model, the Discrepancy model instead of the Provus model, the Antecedent-Transactions-Outcomes model instead of the Stake model, the Context-Inputs-Process-Product model instead of the Stufflebeam model, the Goal Free model instead of the Scriven model number two, and so on. In other words, it is time that the field of evaluation get off of the cult of personalities and get on with an operational consolidation and growth. If we were not so busy competitively trying to enhance our own camp, we might see how the many models compliment each other, thus enhancing knowledge about the evaluation process.

The remainder of this paper displays a synthesis of the extant models of evaluation. It encompasses and allows for the variation in evaluation that has evolved over time. This synthesis is structured by focusing on two dimensions. The first is a continuum which might be labeled, "Degree of Relationship to a Decision," or "A Continuum of Evaluation Models." The second dimension consists of categories of description. Taken together these two dimensions create a three by five matrix as shown in Figure 1. The discussion which follows describes the two dimensions and the categories on each of them. The various evaluation models referred to at the outset are located on Dimension 1; the degree of relationship to a decision continuum. At the same point, the similarities of a number of the extant models are discussed and the argument advanced that basically they are the same model expressed in slightly different language. The next segment of the paper briefly reviews the information for each of the fifteen cells in the matrix; information which needs to be understood in depth by the student of evaluation. The paper closes with some speculation of concerns for the future.

Dimension Number 1: A Continuum of Evaluation Models.

Evaluation in every instance is a purposive activity. In *Educational Evaluation: New Roles, New Means* (Tyler, 1969), a National Society for the Study of Education sponsored group traces the history of evaluation. In all of its stages it is clear that evaluation is a purposive phenomenon. This purpose is illustrated in statements such as:

. . . for critically judging the worth of whatever the schools teach.
. . . to aid instructional planning.
. . . to determine the relative worth of new approaches.
. . . to facilitate the formation of instructional materials and procedures.

Inherent in each of these is the idea of decision making - choice making - the selection of one entity from some set to be treated, used, or considered differently than the remaining members of the set. For example, a student has completed some work that is

to be given a qualitative label. The labels "excellent," "good," "fair," "poor," or "bad" make up that set of possible labels. One of them is to be chosen as an appropriate descriptor of the student's work. Or another example, several textbook series are being considered for teaching reading; one of which is to be selected and the others rejected. The collection of texts is the set of alternatives in this case. Still another example is the case in which certain events are to occur in a sequence of work. At any given point in the sequence there is the possibility of terminating the effort, redoing it just as it was done, redoing it with some minor modification, or redoing it using a totally different approach. The individuals involved must choose one course of action from this set and reject the others. The point to be drawn from these examples is that evaluation's purpose is decision making. Whenever undertaken, evaluation at least facilitates choice making.

Evaluation's relationship to decision making is varied. In some instances the relationship is remote. For example, in the area of psychological services we speak of "evaluating the child." In such a situation tests are administered in the compilation of a case study. It is understood that decisions are to be made about handling the particular child. However, at the time of the evaluation we are not sure of the alternative treatments that may eventually be considered. In other instances the relationship of evaluation to decision making is much more direct. In fact, in some cases the consideration of the relative worth of different alternatives and the selection of one of them are wrapped up in the same events. Examples of evaluation encompassed in decision making can be seen in the accreditation work in education. People engaged in such evaluation simultaneously arrive at judgements and recommendations. Scriven's Goal Free evaluation is another example of this direct tie between evaluation and decision making. These two instances present extremes. On one hand evaluation is an act of measurement which may or may not contribute information helpful in arriving at specific decisions, on the other hand evaluation is an indistinguishable part of the choice making activity leading to judgements and/or recommendations.

Figure 1. A Structure for Synthesizing Models of Evaluation

	Categories of Description	General Statement	*A* *Perhaps Associated with Decisions*	*B* *Isomorphic to a Decision*	*C* *Encompassed in a Decision*
I	General Nature.	Affixing worth statements to entities.	A data banking or measurement process (e.g.; standardized test program).	A service to decision making through which relative worth of alternatives is described.	Making judgments re: the value of entities and making recommendations.
II	Characteristics of "Problem" for which It Is an Appropriate Solution Stragety.	An intention to choose one or some combination from some larger set of alternatives.	Regularized activities in which decisions may be made which might be facilitated through the use of previously generated data	1. Intention to choose a specific alternative. 2. The existence of 2 or more possible alternatives. 3. The lack of data [illegible] estimates relative worth of all alternatives. 4. Lead time.	1. The existence of some completed educational project or an institution. 2. Competing alternatives. 3. Desire or need to control subsequent operations.
III	Process Components or Activities.	Delineation of needed information; Obtaining that information; and Providing the Information.	1. Delineation of the typical decisions. 2. Identification of info which might be used in those decisions. 3. Testing & data storage. 4. Retrieval and interpretation.	1. Determination of the alternatives that will be considered & criteria to be used in weighing them. 2. Generating the information on all alternatives. 3. Communicating info to decision makers.	1. Identifying competent judge(s). 2. Collecting and analyzing data. 3. Reaching conclusions. 4. Submitting recommendations to authorities.

IV	Products or Outcomes of the strategy.	Information regarding relative or absolute worth of entities.	Data on characteristics of people or programs encoded in some bank ready to be accessed if useful in a decision situation.	Information which describes the relative worth of a specific set of alternatives in a particular setting.	Statements of worth and/or recommendations for subsequent actions.
V	Criteria Useful in Assessing the Quality of an Evaluation.	1. Fit between information and decisions. 2. Validity and reliability of data. 3. Accuracy and quality of communication. 4. Cost.	1. Fit between data items & typical decisions. 2. Quality of data generation techniques. 3. Retrievability of stored info. 4. Security of info. 5. Cost.	1. Scientific criteria-validity, reliability. 2. Practical criteria-relevance, importance, scope, credibility, timeliness, pervasiveness. 3. Cost.	1. Credentials of judge. 2. Procedures for data generation. 3. Reportage to people in positions able to carry out recommendations.

Between these two extremes is another degree of relationship between evaluation and decision making. To some people evaluation is a process for providing information useful to decision makers. In this instance, evaluation is directly related to decision making but not the act of choosing. In this form evaluation involves the determination of the nature of the alternatives from which a choice is to be made and the creation of information which describes the relative worth of each of the alternatives. That information is then given to the decision maker or makers who consider the information and then make the actual choice. In this form evaluation is a service to decision making; not the act of decision making as described in one of the extremes above. It differs from the other extreme in the fact that it aims to serve a specific decision.

These three forms of evaluation are pictured in this paper as points along a continuum of degree of relationship between the evaluation activity and decision making. The three points discussed above could be characterized as models of the evaluation process. This dimension is described as a continuum to recognize the possible existence of models between the three points cited above.

Dimension 2: Categories of Description for Evaluation Models

The second element in the proposed structure for synthesizing models of evaluation focuses on ways of describing the different evaluation models. A complex purposive behavior can be described in five ways:

(1) in terms of its general nature,
(2) in terms of the characteristics of problems for which it is an appropriate resolution strategy,
(3) in terms of the activities and events that make up the process itself,
(4) in terms of the products or outcomes achieved once the activity is carried out,
(5) in terms of the criteria useful in assessing the quality of a completed effort.

These five categories have evolved from studies of ways people define complex entities. They are not presented as competing ways of description but rather as complementary ways. Full knowledge about any complex purposive behavior requires an integration of the descriptions from all five categories.

One more point is necessary before dealing with the fifteen cells in the grid presented as Figure 1 in this paper. Evaluation can be described by using terms of differing degrees of abstraction. On one hand it can be described as an aspect of a scientific method. Such a description is inherently abstract. On the other hand, it can be described as the specific data generating activity for a specific decision. In this we would be talking about a specific test or measuring procedure, specified decision makers, subjects to be studied, a particular set of alternatives to be considered, and so on. At the scientific method level of discourse we have a discussion that is grossly abstract, on the other hand it might be extremely specific. As the elements of the grid in Figure 1 are considered, both levels of discourse will be presented. The first level will attempt the general statement that covers all models of evaluation for the category of descriptions being considered. The second level of discourse will look at that category of description as it differentiates for each of the three models of evaluation. To facilitate this discussion the three models will be shorthanded as follows. Model A is an evaluation equals measurement model, in referring to it we will use the phrase "E=M"; evaluation is equated with measurement. The second model is one in which evaluation serves decision making; in referring to it we will use the symbols "E=S", evaluation is a service; an aid to decision makers. Finally, the third model is one in which evaluation is encompassed in decision making or judgement. It will be represented by "E=J". Before getting into the fifteen elements in the matrix, the relationship between existing models and the three points on the continuum proposed needs to be discussed.

Many individuals in the field equate evaluation with measurement. This can be seen in comments by Jack Merwin in the previously mentioned NSSE yearbook. *Educational Evaluation: New Roles and New Means.* Merwin's chapter is entitled,

"Historical Review of Changing Concepts of Evaluation." Most of the content, however, focuses on growth of measurement capabilities in education. There are numerous textbooks which do the same thing. They present a title like "Evaluation and Measurement." An examination of the content of such books shows their equation of evaluation and measurement. These books deal almost exclusively with measurement theory and techniques. The same emphasis can be seen in evaluation help offered by commerical testing companies. To a sizable group of people, then, evaluation is equated with measurement. Recognition of this fact should not be interpreted as a negation of it. There are times in which our work calls for measurement of the phenomena with which we work; measurements which may or may not be used in specific decisions. In other words, there are frequent instances in which we find it valuable to administer a number of measures to an individual or group of individuals so that empirical data will be available when needed.

The proliferation of models of the evaluation process comes at points B and C on the degree of relationship continuum. For a long time individual voices in the field have advocated the use of evaluation in decision making. Ralph Tyler stands tall among those individuals. However, general acceptance of the advice given by Tyler was not great. His encouragement to systematically examine educational activities to provide data for decision making about them was not generally heeded. With the passage of the Elementary and Secondary Education Act of 1965, federal funds provided an incentive for evaluation. In effect, ESEA (1965) added an eleventh commandment for educators; thou shalt evaluate! As a result of that act, many individuals turned their attention to the nature of the evaluation process resulting in the evolution of numerous models of the evaluation process.

Among the leaders in this new rush one can find the names of Michael Scriven, Daniel Stufflebeam, Robert Stake, Malcolm Provus, and Marvin Alkin. Each of these individuals, and others associated with them, have evolved a model of the evaluation process which seems to fit at point B on the continuum.

Scriven's model, the Formative-Summative Model, displays

two general kinds of evaluation. The first, Formative, produces information useful in making decisions in the creation of an educational program or procedure. The second, Summative, generates information related to the accomplishments or output of the program useful in decisions about repeating, terminating, or modifying the program.

The CIPP Model, evolved by Stufflebeam and others, concentrates on four classes of decision making and proposes four types of evaluation activities related to those decision categories. The first category of decision making relates to choice among goals. Context evaluation is proposed to assist decision makers in selecting among competing goals. Inputs evaluation is the process for determining the relative worth of existing plans or strategies for accomplishing a chosen goal. The third category of choice making in this model encompasses those process decisions made while the operation is being carried out. Process evaluation is the strategy for assisting with operational decisions as a project is carried out. The final category of decision making as discussed in this model is recycling decisions. Once you have completed a program the question remains, "What is to be done in the next operational cycle?" Should we make the thing we evaluated a regular part of the program, or try it again with some modification, or give up? Product evaluation is proposed as the CIPP model strategy for generating information useful in decisions of this sort.

The model proposed by Robert Stake, sometimes called the Countenance Model, concentrates on antecedents, transactions, and outcomes. Each of which is to be examined from four vantage points: first, intentions, what are our intended antecedents, transactions, and outcomes; second, observations, what is observed that is related to the three; third, standards, the delineation of standards related to antecedents - transactions outcomes; finally, judgements, what judgements can be made about the antecedents, transactions, and the outcome. The work generated in this three-by-four model provides a store of information useful in decisions about educational programs and products.

Malcolm Provus, in creating the Discrepancy Model of evaluation, also focuses on decision making. The Discrepancy Model starts with the recognition of two entities. The first consists of the standards; the things valued and expected. The second encompasses the actual performances obtained. The idea of discrepancy is inherent and is shown through a comparison of the performances and standards. Such a comparison, according to Provus, is done to provide information about alternative actions. His work identifies three alternatives that are always open: first, the decision to change performance in some manner; second, the decision to change standards; and third, the decision to terminate. Again, the evaluation process is seen as a means of generating data and converting that data to information useful to decision makers.

The model evolved by Alkin and his associates has five major components: need assessment, program p'anning, implementation evaluation, progress evaluation, and outcome evaluation. Materials associated with this model state that evaluation is seen as a process of determining the kinds of decisions that have to be made and selecting, collecting, and interpreting the information needed in making those decisions.

Although these leaders use different terminology in describing their approaches to evaluation, their work is amazingly similar. Several kinds of data support the assertion that all of these are E=S models. First, individuals studying with one of these leaders find little or no difficulty in applying what they have learned when they go to work with another one of the leaders in the field. Second, there have been a few instances in which two or more of these individuals have been asked to work on the same evaluation task. Little or no difficulty is encountered when they work together. In fact, when put to work on the same project their independently developed operational procedures are almost identical. The third point in my argument that these leading theorists have proposed essentially the same model can be seen in the work of Worthen and Hock (1973). These two individuals have prepared a set of simulated materials to facilitate learning about the evaluation process. In the instructor's manual for these simulated materials, Worthen and Hock show the manner in

which the tasks in their simulation relate to the different models of evaluation. In reading through that explanation it is striking how many times the authors are able to say that a specific task relates to the X model the same way it did to the Y model and the Z model. Worthen and Hock make these statements in a way in which few readers question their assertion. Thus, if different models relate to the same task in the same manner, if students studying with the different leaders have little or no problem in orienting themselves when working with one or the other, and if the work performed by the leaders is essentially the same, it seems safe to conclude that despite differences in language or pictorial representations they are proposing the same model. Each of them is an effort to describe rational procedures useful when an individual is assigned the task of generating data and converting it to information about a set of alternatives; information to be used by someone else in making a decision.

The third point on our continuum was described earlier as the evaluation equals judgement model (E=J). Michael Scriven's Goal Free Model fits at this point. Scriven argues that educational activities should be examined by someone unfettered by knowledge of the purposes the program was designed to accomplish and the evolving characteristics of the program. That individual, he suggests, should be free to examine what his dictates tell him to examine and to come up with conclusions about the worth of the work that has been accomplished. Another example of an evaluation model that fits at this point can be seen in the work of accreditation agencies. These agencies gather information and reach judgements regarding the worth of an educational institution. In both of these models the evaluation process and the decision or judgement process are so interwoven that they are one and the same.

In identifying evaluation models at these three points on the continuum two ideas should remain clear. First, the specific location of the model at any point on the continuum is not an attempt to demean the model. It is asserted here that the full continuum of evaluation models is needed for specific purposes in the field of education. If there is anything wrong with being at any one of the three points on this continuum, it is in the

application of the model at that point for a purpose other than the one for which it is specifically developed. Secondly, in showing the similarities among evaluation models, work of some individuals has been overlooked. This is in no way a denegration of their work. Factors of time and space preclude dealing with all the models.

Awareness of this range of models is proposed to facilitate interactions between evaluators and their clients. If the parties involved have different perceptions of the evaluation process, difficulties and disappointments can be expected. The matrix below suggests all the possible combinations.

Evaluation Model Category

Role	E=M	E=S	E=J
Evaluator			
Client			

If the evaluator believes and works from the E=M model, and the client, the decision maker, believes in the E=J model, major disappointments in an evaluative effort are probable. The competent evaluator should: recognize these possible differences; bring them into the open; and resolve them either by agreement on the model to be used or by terminating the interaction.

Description of the Evaluation Models

As indicated earlier, the materials which follow are keyed to the matrix presented as Figure 1. Each segment of the discussion below will be keyed to that matrix through the use of a Roman numeral for each row and a letter for each column. As also stated earlier, two levels of dicussion can apply to each of the rows of matrix. In each the abstract case will be made first and the specific for each column will be present immediately following.

I. General Nature

Evaluation can be described as a process for determining the worth of entities. The use of the term "process" denotes the fact

that evaluation involves a number of component elements which fit together. The concept of worth should be considered in two ways: absolute worth and relative worth. At times in any model of evaluation either of these may be applied. That is, there will be instances in which an effort is made to determine the absolute worth of entities on which the evaluation work focuses. At other times we shall be restricted to or content with relative worth. The choice of concentration on absolute or relative worth is determined by the nature of the alternatives that are involved in a specific decision. Recognition of that calls up a final consideration of the general nature of evaluation. That is, evaluation is situation specific. Unlike research, evaluation does not have generalizability across time and space. In the words of Cronbach and Suppes (1969), evaluation is decision-oriented inquiry in contrast to conclusion-oriented.

The general nature of the evaluation equals measurement model (I A) can be described as a measurement activity. It is the activity engaged in when a quantitative description of the things being worked with is desired. Thus, E=M evaluation model involves the existence of the need to measure and the activity and materials for satisfying that need. This model functions in a situation in which a general class of operations is being carried out and some segment of the workers in that setting see value in having a variety of qualitative and quantitative data about the subjects being worked with.

General nature of the E=S model (I B) is that of a service to decision making. This model applies to situations in which specific decisions which are going to be made at some future point can be identified. It also implies that such decisions can be defined in terms of alternatives and criteria to be used in determining their relative worth. Typically, someone other than the decision makers is assigned the responsibility for generating the data and converting it into information. Evaluation in this model is isomorphic to decision making. That is, the characteristics of the specific decision set the characteristics of the evaluation activity. Every alternative possibly considered in the decision must be represented in the evaluation activity. Further, each criterion to be used in determining the relative value of

alternatives serves as a component in the evaluation activity. The basic purpose of evaluation in this situation is to provide empirically based information to the individual or individuals responsible for making a choice. That individual receives the evaluation report, uses it to examine the different possibilities, and selects a course of action to be taken. In the E=S model evaluation activity leads up to the decision. It may or may not specify the decision which is actually chosen by the decision maker. Thus, the E=S model is a process through which information is provided to the decision makers; information which is hopefully used in guiding choice making.

The E=J model (I C) is one in which evaluation and decision making are thoroughly mixed. Typically, the work of determining relative worth and of arriving at a choice or judgement is done by the same individual. It is almost impossible in the application of such a model to distinguish the elements of either evaluation or judgement making.

II. Characteristics of "Problem" for Which the Model is an Appropriate Solution Strategy

Evaluation is a problem-solving strategy; a purposive strategy appropriate to some classes of problems and inappropriate to others. The term "problem" in this instance implies that an intention exists and that the accomplishment of that intention is blocked in some manner. The problem for which the evaluation strategy is most appropriate can be described as one in which there is an intention to choose. That is, a finite set of possible alternatives exists, one or more of which are to be selected for treatment different than the remaining alternatives in the set. The barrier, the blocking of the accomplishment of that intention that renders it a problem, is the lack of relative worth labels for each of the alternatives, worth labels that are directly appropriate to the situation within which the choice is to be made. This intention-to-choose problem is different from the problem for which the research process is the appropriate strategy. A research problem is one in which we *intend to know* in some general theory sense, an intention which is blocked by a lack of conclusive studies on the item to be known. It also differentiates

from the type of problem for which the development process is a resolution strategy. That problem is one in which it is our *intention to do* some educational task which is blocked by the lack of tools or procedures for doing that task at a satisfactory level of accomplishment. The characteristics of the evaluation problem, then, are an intention to choose among some set of alternatives; an intention which is blocked by the lack of information which relatively weighs the worth of each alternative in the setting in which the choice is to be made.

The problem for the E=M model (II A) is an intention to use quantitative data in operational decisions about individuals. It is a generalized intention which calls for the creation of some general descriptive information. For example, we know that if we are going to eventually make decisions about children's psychological development we will need psychological measures on those children. The problem in the E=S model (II B) is the existence of an intention to make a choice at some point in the future. That intention can be described in terms of alternatives from which the choice will be made. Such a situation becomes a problem when information is not available which describes each of these alternatives on a set of criteria to be used in the choice making. Finally, the problem in the E=J model (II C) is one in which there is an intention to state some qualitative descriptor for a completed program. The purpose for such a statement is to provide guidance for others about that program. This is an instance in which there is an intention to make definitive qualitative statements. Such an intention becomes a problem when it is blocked by the lack of unbiased objective assessments of what was accomplished, both related to and apart from the intended goals of the work.

III. Process Components of Evaluation

Given the general description of evaluation as a strategy for resolving problems of choice making, the processes of that strategy seem, in general, to be: 1) the determination of the items that are to be compared in value, the items from which a choice is to be made; 2) the generation of data necessary in describing the relative worth of these items; and 3) a communication of the

information so generated. Together these activities construct an information base which details the absolute or relative worth of each alternative. That information base is made available or provided to the individual making the choice among the alternatives.

The process components of the evaluation equals measurement model (III A) include: delineation of the general set of descriptive parameters on which quantification should take place; generating data descriptive of each unit of the population on each of the parameters; and storage of those data in a manner that is both easily accessed and secure. The evaluator applying this model determines what categories of information are generally useful in decisions likely to be made about the subjects in this setting. The set of parameters he identifies should be checked for accuracy and adequacy with those people working in the environment and with individuals who have expertise related to that type of setting. The delineation activity is complete when a set of measures have been identified which are agreed by the participants, the decision makers in the setting, to have general utility for a variety of decisions to be made about the subjects. The second activity in the E=M model is the generation of data on the subjects in the population. The third activity involves the storing of the information generated in an easily retrievable form. When decisions are to be made about any individual in the population, those persons responsible for the decisions must have easy access to the information. At the same time, the security factor related to the data must be attended to.

The activities of the E=S model (III B) start with the delineation of the information needs for the specific decision to be served. That delineation starts with the specification of the alternatives to be considered by the decision makers and the criteria to be used in that consideration. The evaluator should solicit from the decision maker the alternatives with which he is familiar for the given situation. The list should be augmented by reviewing literature, by conversations with consultants, and by study of other sources of information about possible alternatives and criteria. The delineation activity specifies a matrix of information needed. Each column in the matrix is labeled by a

criterion variable to be used in relatively weighing the different alternatives. Each row in the matrix is labeled by an alternative to be considered. The second activity in the evaluation process under the E=S model is obtaining the information for each of the cells in that matrix. A variety of types of data generation techniques may need to be applied to generate the information needed in a specific study. The third activity in this model is the provision of information to the decision maker or makers. An important ingredient here is the manner in which the decision maker receives information most efficiently. Once that is determined, guidance for the reporting or information provision activity is in hand.

The activities of the E=J model (III C) involve: the determination of the outcomes of a project, both positive and negative; determining the relationship of those outcomes to the intentions of the project; and finally, creating judgement statements regarding the quality of the outcomes and the worth of the effort along with recommendations about it or its future.

IV. Products of the Evaluation Strategy

The principal outcome of evaluation is empirically-based information rationally related to choice making. The information produced is in the form of worth labels that are helpful in determining the relative value of the different alternatives in some set. For example, a set of alternatives might be four ways of beginning reading instruction. An evaluation effort related to these beginning reading instruction alternatives would produce value labels to be attached to each of the four alternatives. These value labels ought to aid an individual in determining which alternative is best for his situation. Thus, the direct product of evaluation is information that symbolizes the value of some entities in a specified situation.

In the evaluation equals measurement model (IV A) these products are numbers associated with specific subjects. Numbers which symbolize different characteristics of the subjects and the relationships of those characteristics to other individuals in the population as a whole. The product of the E=S model (IV B) is

information which describes relative worth of each of the alternatives considered in the specific situation. That relative worth is shown on each of the variables to be used by decision makers as they make choices among the alternatives. The product in this case is informational. It has not moved to the finality of a decision as is the case in the evaluation equals judgement model. Here the product (IV C) exists in the form of qualitative descriptions and recommendations for action.

V. Criteria Useful in Assessing the Quality of an Evaluation

The criteria useful in assessing the quality of an evaluation effort are the scientific quality of the information generated and the general usefulness of the information in the decisions to be made. In the first of these, the quality of the information generated, we are interested in whether the real world is accurately pictured by the information generated. If an evaluation report says that one alternative will yield so much educational benefit and a second alternative a little bit less, is this an accurate description of the alternatives considered? Central to this concern for evaluation quality are the criteria typically applied to any scientific or empirically-based effort; the practical criteria for the assessment of an evaluation's quality focus on the utility of the information in the decision-making process. Here we are concerned about directness of relationship to the decision, credibility of the information, inclusiveness, timeliness, etc.

In the evaluation equals measurement model these criteria (V A) make up the following concerns: 1) does the set of parameters to be used in the evaluation yield information related to the major types of work undertaken in that setting? (This is a comprehensiveness criterion. Are all of the necessary decision making activities going to be covered uniformally in the measurement work?); 2) does the information collected actually contribute to the making of decisions? 3) are the measuring procedures in fact valid and reliable? 4) is the information stored in a manner in which people who are going to be making decisions can readily access the information - is it available yet under sufficient security to protect the rights of the individuals in the setting? 5) is the cost of the activities justified by the magnitude of the

decisions to which these data are related?

In the E=S model of evaluation, three general sets of criteria need to be considered in assessing the quality of a completed evaluation effort (V B) (Stufflebeam, et al. 1971). The first set includes the scientific criteria. Here the concern is for the degree of validity of the information, the degree of generalizability, and the degree of objectivity. The second category includes the practical criteria of relevance, importance, scope, credibility, timeliness, and pervasiveness. These concerns deal with the degree to which the evaluation data is relevant to the decisions to be made. The criterion of importance asks: are the trivial items of information eliminated? Scope deals with the degree to which the full range of elements of the decisions are treated. Credibility takes a slightly different focus. Here the concern is for the degree of trust or belief that is inherent in the information. Timeliness takes still a different focus. In this area one is concerned with the degree to which the information arrives in advance to the actual decision. Finally, pervasiveness is the degree to which the information gets to all of the decision makers. The third category of criteria for the E=S evaluation model is again, cost. An evaluation activity needs to be cost efficient.

Criteria useful in assessing the quality of an evaluation done under the evaluation equals judgement model (V C) take the following form. First, are the credentials of the evaluator sufficient to merit trust in the judgements he will be expected to make? Second, are the procedures used in the investigation for the collection of information of acceptable quality? Third, does the study conducted meet standard investigative criteria? Fourth, are the judgements delivered to those audiences in a position to do something about them? Finally, a criteria of cost must be considered. Again, the bill for conducting an evaluation equals judgement effort ought to be relative to the cost of the overall project being evaluated.

THE PAST AND FUTURE OF EVALUATION

The process of evaluation has evolved over this century. Its roots come from three places. First, the educational measurement

movement. As measurement technology advanced and educators became more and more capable of providing quantitative descriptions of educational phenomena, our ability to evaluate has improved. The second element is the growth of scientific research in education. By evolving methods and techniques for enhancing knowledge about learning and educational procedures related to it, evaluation became a more viable activity. The third force leading to evaluation in all of its forms today is the accountability movement. As the American society grew more and more concerned about the value of the education program it was purchasing, greater emphasis was gained for engaging in evaluation activites. Each of these sources can be seen in aspects of evaluation processes that exist today. Measurement is inherent in any form of evaluation. Sometimes the measurement involves carefully developed instruments, in other cases some subjective impressions are the form in which measurement occurs. Regardless of the form, however, measurement is always involved. Many elements of scientific research including sampling, experimentation, and data analysis can be seen in evaluation work. In fact, most of evaluation's technology is borrowed directly from research. The accountability movement and its concern for efficiency can be seen in most evaluation work. We are striving for rational improvement and looking to evaluation as one means for its accomplishment.

The bulk of this paper details the present by describing the nature of evaluation in its many current forms. As guidance in considering evaluation's future, a different view of evaluation ought to be taken. What is the psychological nature of the environment in which evaluation exists today? Phi Delta Kappa's National Study Committee on Evaluation took a look at this a few years ago. That view led to the conclusion that evaluation is sick. In reviewing statements made by that group in the book, *Educational Evaluation and Decision Making,* it is this writer's belief that most of the symptoms of evaluation's illness still apply today. What are those symptoms? First, we have an avoidance symptom. Evaluation seems to be a painful process; one which everyone avoids unless it is absolutely necessary. Educators in particular avoid it. Evidence supporting of that assertion can be obtained by examining the evaluation component of any education budget. If money is budgeted, it is minimal! A second

symptom of evaluation's illness is anxiety. People who engage in evaluation have anxieties about the work they are to do, and individuals who are subjects of evaluations are even more anxious. The evaluator is concerned because he knows how inadequate and subject to error the evaluation process is. The client is anxious because of his lack of ability to predict what might happen. The third symptom cited by the National Study Committee was immobilization. It is still obvious today. Most educational operations have not implemented evaluation as an ongoing part of the operation nor as an ad hoc activity related to the decisions to be made. The fourth symptom was cited as skepticism. Many persons seem to argue that there is little point in planning for evaluation because it can't be done anyway. This is in part an outgrowth to the diversity in models of evaluation discussed earlier in this paper. If the experts do not agree about the nature of evaluation, how can anyone else be expected to effectively carry it out? The next symptom of evaluation's illness the National Study Committee called a lack-of-guidelines symptom. This shows itself in relation to situations where evaluation is mandated. Seldom does that mandate also provide effective guidelines for carrying out the evaluation that is required. The misadvice symptom, the sixth listed by the Committee, is a difficult one to handle. Many instances can be cited in which "experts" have recommended evaluation activities that are plainly misadvice. Examples may be seen in the recommendations by researchers which implicitly equate evaluation and research. Such individuals impose activities that do not necessarily facilitate decision making. Misadvice also can be found in those instances in which a group is advised to turn to some other discipline for the evaluation effort. Too often the "expert" from the other discipline simply applies research techniques generally appropriate in that discipline. Such instances also typically fail to provide the information which aids decision making related to educational programs. The next symptom identified by the National Committee was labeled as the no-significant-difference symptom. Examination of evaluation reports frequently show a statement that significant differences were not observed. At the same time, people in the setting using the different methods can point to marked changes related to one or more of the procedures being tried. Evaluation's failure to identify that kind of information is

clearly an indication of its illness. The final symptom of evaluation's illness was cited by the National Study Committee as a missing-elements symptom. This is a focus on the evaluation capabilities in the field of education itself. We lack comprehensive theory related to evaluation. We lack a clear specification of the kind of evaluational information that would be most useful. We lack instruments designed for conducting evaluation; each evaluation starts as a new single effort. We need mechanisms for organizing, processing, and reporting evaluative information. Finally, the most important missing ingredient is that of trained personnel. We just do not have enough competent people to do the necessary work.

Taken together these eight symptoms provide a target for the work that needs to be done *if* evaluation is to be cured. Efforts to cure this illness, however, are not strongly visible. The field is fragmented. As noted earlier, we're suffering from a personality cult. We have camps among us trying sincerely to improve *their* model of evaluation. We need to merge these efforts if we are going to successfully attack these eight symptoms. Accomplishing that merger is going to be extremely difficult for no agency currently exists with the strength and desire to help the field of educational evaluation set goals and evolve the mechanisms for accomplishing those goals. The future looks bleak. Unless change forces can be produced, a possibility of attacking successfully and eliminating the symptoms of evaluation is out of the question. Three general types of change force are needed. The first is a concentration on the conceptual base we have for evaluation. We need a unifying theory. Individuals with both abstract and practical orientation need to attend to this change. A second focus for change is that of developing the interface between evaluation and other components of the education system. It is not enough to recognize that administrators do not understand what evaluation can do for them. It is not enough to recognize the teachers are anxious about being evaluated or that they don't know how to use the information. Work on evaluation must be advanced in a way that makes it meaningful to the other elements of the education system; a way which improves the interface. The final focus for the change is in the training of evaluation personnel. We must provide the manpower that understands the

nature of the evaluation process and the modifications needed to be a practical useful service in any given institution. A number of institutions are interested in preparing this manpower. At present, the capability of most of them is very low. We have to assist these interested institutions in doing the quality education jobs needed for the preparation of competent evaluation personnel. Attention to all three of these forces for change is the challenge of the future.

OCCASIONAL PAPERS

1. THE PROBLEM AND PROBLEM DELINEATION TECHNIQUES – William J. Gephart, Phi Delta Kappa. Presented at the Second National Symposium for Professors of Educational Research, sponsored by Phi Delta Kappa, Boulder, Colorado, November 21, 1968. A discussion of the nature of the concept "problem" as related to educational research with a discussion of several techniques useful in problem identification and delineation. $1.00
2. A REVIEW OF INSTRUMENTS DEVELOPED TO BE USED IN THE EVALUATION OF THE ADEQUACY OF REPORTED RESEARCH – Bruce B. Bartos, Phi Delta Kappa & Indiana University. Presented at the Annual Meeting of the American Educational Research Association, February, 1969, Los Angeles, California. A brief description and bibliographic annotation of 40 instruments developed to be used in assessing the methodological quality of completed research. $.25
3. PROFILING EDUCATIONAL RESEARCH – William J. Gephart, Phi Delta Kappa, January, 1969. The rationale for the development of a methodology profile on completed research to show its strengths and weaknesses. Included are flow charts for profiling the five facets of the research process. $.75
4. APPLICATION OF THE CONVERGENCE TECHNIQUE TO READING – William J. Gephart, Phi Delta Kappa, January, 1969. An interim report on a research program planning effort in the field of reading. (Out of Print)
5. THE CONVERGENCE TECHNIQUE AND READING: A PROGRESS REPORT – William J. Gephart, Phi Delta Kappa. Presented at the Annual Meeting of the International Reading Association, May 2, 1969, Kansas City, Missouri. A second interim report on the planning of a reading research program. (Out of Print)
6. THE EIGHT GENERAL RESEARCH METHODOLOGIES: A FACET ANALYSIS OF THE RESEARCH PROCESS – William J. Gephart, Phi Delta Kappa, July 14, 1969. The identification and description of general research methods in education through the use of Gutman's facet design and analysis technique. It also details the procedures for the Gutman technique. This paper was printed in the proceedings of the Warsaw, Poland Congress of the International Association for the Advancement of Educational Research. $1.00
7. PROFILING INSTRUCTIONAL PACKAGE – William J. Gephart & Bruce B. Bartos, Phi Delta Kappa, August, 1969. An instruction text to assist individuals with no prior research training in the use of research profiling flow charts to assess the methodological adequacy of completed research. $1.00
8. EDUCATIONAL KNOWLEDGE USE – Gene V. Glass, Laboratory of Educational Research, University of Colorado. An analysis of the availability and use of empirically based information in education. $.50
9. MEASUREMENT AND RESEARCH IN THE SERVICE OF EDUCATION – Warren G. Findley, Research and Development Center in Educational Stimulation, University of Georgia. Originally presented as an invited address at the annual meeting of the American Educational Research Association; this paper uses an historical perspective to examine the role of measurement and research in education. $.75
10. THE EDUCATIONAL CATALYST: AN IMPERATIVE FOR TODAY – Joe H. Ward, Jr., Reeve Love, and George M. Higginson, Southwest

Educational Development Laboratory, Austin, Texas, July, 1971. An analysis of the problems involved in the process of change and improvement of the practice of education. This paper poses a new professional speciality for the facilitation of empirically based educational improvements. $1.00

11. DISSERTATIONS YOU MAY WANT TO SEE – William J. Gephart, Phi Delta Kappa, 1970. A collection of dissertations done in 1969 which focus on research training. $.25
12. THE DOCTORATE IN EDUCATION IN CANADA – Neville L. Robertson, Commission on Higher Education, Phi Delta Kappa, 1971. An analysis of the institutions offering the doctorate in education in Canada. This paper is a companion piece to a larger study of similar institutions in the United States. $.75
13. THE IMPORTANCE OF STATISTICAL POWER IN EDUCATIONAL RESEARCH – John K. Miller, and Thomas R. Knapp, University of Rochester. When an educational experiment results in non-significant differences can it be said that no difference exists? This paper discusses the concept that must be attended to IF that question is to be answered. It also details the procedure for determination of sample size needed in an experiment. $1.25
14. THE CONDUCT OF DEVELOPMENT IN EDUCATION – Richard E. Schutz, SWRL Educational Research and Development. The past decade has developed awareness of the tools and procedures needed to perform educational tasks. Some have argued that the research process is the vehicle for achieving that end. Schutz argues that the development process is different from the research process in many important ways that need to be understood by educators. $1.50
15. THE PROCESS OF DEVELOPMENT – Gerald Nadler and William J. Gephart. The development process is examined from five points of view in this interview of Gerald Nadler, Chairman of the Industrial Engineering Department, University of Wisconsin. Questions posed by Gephart and their discussion examine the general nature of the process, the types of problems for which it is an appropriate resolution strategy, the component activities of the development process, the types of products resulting from it, and the criteria that would be used in assessing the quality of a development effort. $1.50
16. THE MOST SIGNIFICANT EDUCATIONAL RESEARCH CONTRIBUTIONS OF THE PAST TEN YEARS – William J. Gephart, Phi Delta Kappa, in collaboration with Berlie Fallon, Anton J. Netusil, and William K. Poston, Jr. Three individuals were asked to identify important research contributions. Work thus identified is briefly summarized and comments are made about the manner in which educational research *ought* to be looked to by educators. $1.25

REFERENCES

Alkin, Marvin C., *Evaluation Comment.* 2: 2-7, 1969. Center for the Study of Evaluation, Los Angeles, California.

Cronbach, Lee J. and Suppes, Patrick, (Eds.), *Research for Tomorrow's Schools: Disciplined Inquiry for Education.* Toronto, Ontario: The MacMillian Company, 1969.

Klein, Stephen P., et al. *Evaluation Workshop I.* Center for the Study of Evaluation UCLA, Monterey, California: CTB/McGraw Hill, 1971.

Provus, Malcolm, *Discrepancy Evaluation.* Berkeley, California: McCutchan Publishing Company. 1971.

Stake, Robert E., "The Countenance of Educational Evaluation." *Teachers College Record* 68: 523-540; 1967.

Stufflebeam, Daniel L., et al., *Educational Evaluation and Decision Making.* Itasca, Illinois: F. E. Peacock Publishers, Inc., 1971.

Tyler, Ralph W., (Ed.) *Educational Evaluation: New Roles, New Means.* The Sixty-eighth Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press. 1969.

Worthen, Blaine R. and Michael Hock, (Instructors Manual for Evaluation Simulation.) Worthington, Ohio: Charles A. Jones Publishing Company. (In press.)

Worthen, Blaine R. and Sanders, James R. (Eds.) *Educational Evaluation: Theory and Practice.* Worthington, Ohio: Charles A. Jones Publishing Company. 1972.